Environment International 92–93 (2016) 585–589

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Preface

Using GRADE to respond to health questions with different levels of urgency

Kristina A. Thayer [a], Holger J. Schünemann [b,*]

[a] Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, P.O. Box 12233, Mail Drop K2-02, Research Triangle Park, NC 27709, USA
[b] Department of Clinical Epidemiology & Biostatistics, Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

article info

Article history:
Received 15 March 2016
Received in revised form 21 March 2016
Accepted 21 March 2016
Available online 26 April 2016

abstract

Increasing interest exists in applying the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to environmental health evidence. While ideally applied to evidence synthesized in systematic reviews and corresponding summary tables, such as evidence profiles, GRADE's correct application requires that "the evidence that was assessed and the methods that were used to identify and appraise that evidence should be clearly described." In this article, we suggest that GRADE could be applied to evidence assembled from narrative reviews, modelled (indirect) evidence, or evidence assembled as part of a rapid response, if the underlying judgments about the certainty in this evidence are based on the relevant GRADE domains and provided transparently. Health questions that require assessing the certainty in a body of evidence to provide trustworthy answers may range from hours, to days or weeks, to a few months to scenarios that allow assessing evidence without short-term time pressures. Time frames of emergent, urgent or rapid evidence assessments will often require relying on existing summaries or rapidly compiling the available evidence and making assessments. Even without available full systematic reviews, expressing the certainty in the evidence can provide useful guidance for users of the evidence and those who evaluate certainty in effects. The ratings also help clarifying disagreement between organizations tackling similar questions about the evidence. Using the structured GRADE domains, narrative or other summaries of the evidence can be presented transparently.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Evident from this special issue of Environment International, increasing interest exists in applying the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to environmental health evidence (Morgan et al., 2016; Rooney et al., 2016). GRADE provides a structured framework for assessing certainty, or confidence, in a body of evidence and supporting decisions (Guyatt et al., 2011a). However, misperceptions exist that GRADE is only applicable to clinical questions, full systematic reviews and requires too much time and effort to apply its principles to other evidence summaries. In this editorial we suggest that GRADE can be applied in a variety of decision-making contexts, including urgent responses. While ideally applied to evidence synthesized in systematic reviews and corresponding summary tables, such as evidence profiles, GRADE's application requires that "the evidence that was assessed and the methods that were used to identify and appraise that evidence should be clearly described" (GRADE_Working_Group). Thus, GRADE provides a flexible framework for appraising evidence that could be applied to evidence assembled from narrative reviews, modelled (indirect) evidence, or evidence assembled as part of a rapid response, if the underlying judgments about the certainty in this evidence are based on the relevant GRADE domains and transparently described. The objective of this article is to provide information on how GRADE could be utilized to evaluate bodies of evidence from emergencies to routine evaluations. We focus on GRADE's "certainty in the evidence" domain as it may apply to interventions and risk assessment (Balshem et al., 2011; Iorio et al., 2015). Our examples include assessments of environmental risks and interventions for infectious diseases. We will not address the issue that GRADE has been applied widely to non-clinical questions such as public health and health policy because this has been done elsewhere (Harder et al., 2014; Burford et al., 2012).

2. How GRADE could be applied to scenarios with different time lines

Health questions that require assessing the certainty in a body of evidence to provide trustworthy answers may range from hours, to days or weeks, to a few months to scenarios that allow assessing evidence without short-term time pressures. Based on succinct examples (Table 1), we will describe how applying GRADE provides useful guidance for assessing evidence across these timeframes. We emphasize that GRADE's certainty in evidence domains risk of bias,

* Corresponding author.
E-mail addresses: thayer@niehs.nih.gov (K.A. Thayer), schuneh@mcmaster.ca (H.J. Schünemann).

http://dx.doi.org/10.1016/j.envint.2016.03.027
0160-4120/© 2016 Elsevier Ltd. All rights reserved.

EPA-HQ-2018-0008760045886

Table 1
Examples of GRADE applied across different time scenarios.

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Example	West Virginia Elk River spill Population: community exposed to the chemical spill. Intervention/exposure: chemicals in the spill that contaminated water supply. Comparison: no chemicals in the spill. Outcomes: genotoxicity, developmental or reproductive toxicity, liver toxicity and others.	Melamine in composite food products Population: healthy people Intervention/exposure: melamine from composition food products below 0.5 mg/kg body weight per day. Comparison: higher than 0.5 mg/kg body weight of melamine from composition food. Outcomes: renal insufficiency (assessed with renal clearance), urinary tract calculi, urinary tumors (used for this example of the certainty in the evidence).	Avian influenza Population: people with suspected avian influenza infection. Intervention/exposure: oseltamivir. Comparison: no oseltamivir. Outcomes: mortality, duration of hospitalization, incidence of lower respiratory tract complications (used for this example of the certainty assessment below), antiviral drug resistance existing before treatment, and serious adverse events.	PFOA and birth weight Population: women of reproductive age and fetuses (before and/or during pregnancy or development). Intervention/exposure: perfluorooctanoic acid (PFOA; CAS# 335-67-1) or its salts. Comparison: lower levels of PFOA. Outcomes: fetal growth, birth weight, other measures of fetal or newborn size.
Type of evidence	Available evidence: animal toxicology studies in rodents for two chemicals in the spill (a 28-day study and a teratology study) and SAR analyses for other chemicals in the spill with no toxicology data.	Available evidence: animal toxicology studies in rat and mice with exposures to various levels of melamine via feeding, including a control group. The utilizedevidence should be supported by a literature search with transparent inclusion and exclusion criteria and a (narrative) summary of that evidence.	Available evidence: five randomized trials in patients with seasonal flu (summarized in systematic reviews), case studies of patients with avian influenza, in vitro and in vivo animal data.	Available evidence: a systematic review of 18 non-randomized (observational) studies (10 were included in a meta-analysis).
GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judgments (note these are not necessarily reflecting judgments in the original scenarios).				
Risk of bias	Animalstudies: would be assessedby risk of bias (RoB) considerations for animal studies (e.g. randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals were accounted for). Ideally, RoB assessments would be available for individual studies and summarized across studies. In the Elk River example, the number of animal studies was small and could be assessed at the individual level within a short-time frame. A de novo risk of bias evaluation may not be feasible in cases where evidence is drawn from existing narrative risk assessments that summarize a large body of literature. Nevertheless, it may still be possible to assess risk of bias based on the uncertainties and evidence limitations described in the risk assessment. SAR: could be assessed using OECD model validation or similar guidance that recommends presentation of a defined domain of applicability for a defined endpoint supported by appropriate measures of goodness-of-fit (OECD, 2007).	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, pathologists blinded in their assessments or all animals accounted for). In this case it appears that the animal studies did not report that it was randomized and, thus, may be at risk of bias.	Not serious	Serious based on some concern of risk of bias in the included studies (in the original report, the authors used an approach to rating certainty that accounted for risk of bias by lowering the certainty from high to moderate).
Imprecision	Could be assessed for both animal data and SAR (e.g., considering statistical or numerical uncertainty in model parameters).	While no summary estimates are available, an assessment could be guided by the availability of data from only 100 animals in different exposure groups which would result in wide confidence intervals.	Serious	Not serious
Inconsistency	Could be assessed for both animal data and SAR (e.g., assessing similarity of results based on applying different models).	Only one study was included and therefore no inconsistency is present (Guyatt et al., 2011d).	Not serious	Not serious
Publication bias	Could be assessed for both animal studies and SAR. A judgment of undetected might be reasonable if emergency consultation with scientists reveals that they are not aware of unpublished data (this is to increase transparency of judgments but it would not be the ideal way of addressing publication bias).	Could be assessed using guidance for animal studies but a judgment of undetected might be reasonable if consultation with scientists reveals that they are not aware of unpublished data (this is to increase transparency of judgments but will not be the ideal way of addressing publication bias).	Undetected	Undetected

EPA-HQ-2018-0008760045887

Table 1 (continued)

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Indirectness	Animal studies: could be assessed using GRADE's indirectness assessment (Guyatt et al., 2011c; Schünemann et al., 2013). Animal studies may be rated down for indirectness if concerns exist about extrapolating from animals to humans, e.g., relevance of animal model for the health outcome of interest or route of exposure. SAR: could be assessed based on evidence of direct relation of the model to a defined endpoint. SAR would typically be downgraded for indirectness.	This could be rated down for serious indirectness of extrapolating from animals to humans and uncertainty about the levels of exposure (different levels or routes of exposure evaluated than those one is interested in and modeling of exposure levels based on composition food products from more exact exposures fed to animals). Further concerns would likely be described for the comparator.	Very serious	Not serious
Magnitude of effect	Could be evaluated for animal studies. It may also be possible to assess in SAR in cases where SAR results allow linkage to a data rich chemical. Would not be applicable if there were concerns for risk of bias (Guyatt et al., 2011f).	Could be assessed comparing different exposure levels and observe if tumor rates increase by factors of more than 2 fold across exposure levels if there was no serious concern about risk of bias and if there was no important imprecision (Iorio et al., 2015).	Not present	The effects were not large.
Opposing plausible residual bias and confounding	Could be evaluated as in other non-randomized studies but not present in the current example.	Not present	Not present	There was no evidence of those possible residual confounders or biases would reduce effect estimate.
Dose effect	Could be evaluated for animal studies. May be possible to assess in SAR in cases where SAR results allow linkage to a data rich chemical.	Not present [the data showed a rate of transitional-cell carcinomas in the urinary bladder of male rats of the following rate: controls, 0/45; low dose, 0/50; high-dose, 8/49 (16%)]. No increase was observed in female rats.	Not present	Several studies in which association was modeled by categorized incremental exposure showed evidence of a dose–response relationship, but given that there was no downgrading this domain would not be considered relevant.
Certainty in the evidence for the outcome across the evaluated evidence	The GRADE approach for interventions would be used. Various reasons for downgrading may apply to this example for the outcome of interest. Other scenarios may not exhibit these limitations.	The GRADE approach for interventions would be used. Various reasons for downgrading would likely apply to this example (including indirectness) that would lower the certainty in the evidence for the outcome of interest.	The GRADE approach for interventions was used. Given several reasons for downgrading, the certainty in the evidence for this outcome was considered very low (the lowest of four categories in GRADE).	The authors of the original report used an approach for interventions with starting observational studies as "moderate" certainty in the evidence. However, the GRADE approach for prognostics studies would be used in for a risk assessment in which observational studies are the proper study design and start with "high" certainty in the evidence (Iorio et al., 2015; Spencer et al., 2012). Thus, for this type of exposure studies, the evidence rating may be high for the association between the exposure and the outcome.
Possible summary statement[□]	There is low certainty in the evidence suggesting no association between the exposure and toxicity based on SAR analyses.	There is very low certainty in the evidence suggesting no association between levels of melamine exposure from composition food products below 0.5 mg/kg body weight per day and urinary tumors.	There is very low certainty suggesting that oseltamivir reduces hospitalization in patients with avian influenza.	There is moderate certainty in the evidence suggesting that PFOA is associated with harmful effects on fetal growth.

□ Note, this hypothetical summary was derived by the authors of this editorial, not those of the original report.

indirectness, imprecision, inconsistency, publication bias, large effects, opposing plausible residual bias and confounding and dose effects should be considered regardless of the time available or format of the accumulated evidence (Balshem et al., 2011; Guyatt et al., 2011b, 2011c, 2011d, 2011e, 2011f, 2011g). We believe this can be achieved with varying levels of detail and put emphasis on transparency of the assessment and the description of the approach used to compile the evidence. Based on succinct examples (Table 1), we will describe how applying GRADE provides useful guidance for assessing evidence across these timeframes. We emphasize that GRADE's certainty in evidence domains risk of bias, indirectness, imprecision, inconsistency, publication bias, large effects, opposing plausible residual bias and confounding and dose effects should be considered regardless of the time available or format of the accumulated evidence (Balshem et al., 2011; Guyatt et al., 2011b, 2011c, 2011d, 2011e, 2011f, 2011g). We believe this can be achieved with varying levels of detail and put emphasis on transparency of the assessment and the description of the approach used to compile the evidence.

2.1. GRADE in emergency responses or ultra short time frames: one or more hours

Environmental exposures to potentially harmful chemicals often require extremely fast responses. For data-rich chemicals with a large literature base, the evidence required to assess potenti concerns has likely already been assembled in existing risk assessment(s). For example, in the Flint drinking water lead-contamination crisis attention immediately focused on strategies to mitigate exposure (EPA, 2016; U.S. Department of Health and Human Services). Such emergencies will benefit from existing evidence assessments that express a level of certainty in the evidence.

EPA-HQ-2018-0008760045888

In other emergencies, chemicals may be involved with no or only a limited number of human (likely observational) or animal studies available. Increasingly, evidence on such data-poor chemicals is derived from high throughput screening (HTS) data or structure activity relationship (SAR) analyses to assess the relationship between a chemical's molecular structure and its potential biological activity. The West Virginia Elk River chemical spill illustrates this scenario. In January 2014, 10,000 gal of a liquid used to wash coal and remove impurities that contribute to pollution during combustion were spilled from a leaking tank into the West Virginia Elk River. The water supply of nearly 300,000 people in the Charleston, West Virginia metropolitan area was contaminated. The National Toxicology Program (NTP) conducted urgent SAR analyses on chemical constituents of the spill for which no toxicological data was available (National Toxicology Program, 2014). Results suggested chemicals in the spill were of limited toxicological concern. Aspects of the GRADE domains (using alternative terminology) describing the certainty in the evidence were considered when evaluating confidence in the data. In fact, all GRADE domains could have been used to provide a structured approach to assessing the certainty in the evidence (Table 1). For instance, indirectness may reduce certainty because the direct application of SAR predictions is not well-characterized for the outcomes of interest, especially for non-genotoxic outcomes. The Organization for Economic Co-operation and Development (OECD) principles for considering a SAR model for regulatory purposes use criteria that resemble GRADE domains, including presentation of a defined domain of applicability (indirectness) that may be used for rating the certainty according to GRADE (OECD, 2007).

2.2. GRADE in urgent responses: one to two weeks

On September 19, 2008, the European Commission requested that the European Food Safety Authority (EFSA) provides urgent scientific advice within five days, on the risks to human health due to the possible presence of melamine in composite food products (EFSA, 2008). EFSA utilized its own 2007 report that evaluated the toxic effects of melamine on exposed animals (EFSA, 2007). The original report, based on an extensive literature search, referred to different levels of exposures of rats and mice to melamine and occurrence of urinary tract calculi and malignancies, renal function and other outcomes. That report ends with a recommendation for a tolerable daily intake (TDI). In the 2008 report, EFSA evaluated the health risks of melamine in imported infant formula and composite foods containing melamine. The evaluated scenarios showed that "estimated exposure does not raise concerns for the health of adults in Europe should they consume chocolates and biscuits containing contaminated milk powder." This assessment could have been complemented by a certainty statement based on the GRADE domains (risk of bias: randomized versus non-randomized exposure studies and risk of bias criteria specific for the study designs; inconsistency: did studies show similar results when they should?; imprecision: how precise were the confidence intervals or how many animals were involved in the studies?; indirectness: do animal data, levels of melamine measured, outcomes assessed represent the health question at hand; publication bias: were there unpublished studies that would have altered the results?). If there was uncertainty related to the evidence in the 2007 report (potentially as a result of indirectness when extrapolating from animals to humans and to human exposure from the dose levels tested), expressing the degree of certainty transparently would benefit subsequent assessments. A rating according to GRADE could also inform if the indirectness of melamine exposure from composite food or formula leaves sufficient uncertainty to re-evaluate the evidence when more data is available. Table 1 provides a guide for how the certainty in this evidence could have been rated in a short time-frame. Thus, although not a complete systematic review, a structured rating of the GRADE domains is possible and perhaps desirable to provide transparency about the judgments that are inevitably involved. Furthermore, regardless of the level of certainty in the evidence, given the harmful effects and little or no expected benefits from contamination with melamine, an agency could express a strong recommendation for low levels of acceptable exposure to melamine (Neumann et al., 2016).

2.3. GRADE in rapid responses: one to three months

In 2006, the World Health Organization (WHO) responded to an international fear of human infection with avian influenza virus H5N1. Considered interventions to treat the condition included pharmacological treatment with antivirals, but direct evidence about treatment effects with antivirals for H5N1 infection was restricted to a limited number of cases. Thus, WHO relied on indirect evidence from other influenza treatment data when it evaluated the certainty in the evidence (called quality of evidence in the WHO report in agreement with GRADE's original terminology) of the effects of interventions on various patient important outcomes (Schunemann et al., 2007a, 2007b). The process included identifying existing systematic reviews of indirect evidence for the use of antivirals in patients with common influenza virus infection, supplemented by searches for recent research studies including human studies, animal and in vitro studies. In vitro and surveillance data obtained in the laboratory, for example, influenced judgments about the anticipated resistance to some of the considered pharmacological agents. Within approximately two months, a technical team prepared GRADE full evidence profiles that included a rating of the certainty in the evidence. The certainty was lowered for very serious indirectness and imprecision for several outcomes, including mortality (Table 1). This assessment allowed issuing evidence-based recommendations three months after the decision to develop WHO guidelines was made.

2.4. GRADE for routine responses: more than 3 months

GRADE has been used for routine assessments of evidence taking months to over a year to complete. One of the first applications of GRADE in environmental health was an analysis by the Navigation Guide of the effect of perfluorooctanoic acid (PFOA) exposure on birth weight (Lam et al., 2014). Evidence from humans and animals was separately compiled in two systematic reviews and evaluated using the GRADE certainty in the evidence domain (Johnson et al., 2014). Table 1 summarizes how GRADE domains were used to evaluate the certainty in the evidence in the relevant human studies. After integrating results from the animal and human studies, the authors concluded that PFOA was "known to be toxic."

3. Summary

The GRADE approach can be useful to assess the certainty or confidence in evidence of an association or intervention effect. It requires a structured but not necessarily time-consuming assessment according to the eight GRADE domains. While systematic reviews and meta-analysis facilitate and enhance the credibility of this assessment and can be done rapidly (Schunemann and Moja, 2015), they often are time-consuming aspects. Because of time and resource constraints in specific situations, they are not a "sine no qua" requirement to apply GRADE (GRADE_Working_Group), in particular under time pressure. Time frames of emergent, urgent or rapid evidence assessments will often require relying on existing summaries or rapidly compiling the available evidence and making assessments. Even without a full systematic review, expressing the certainty in the evidence can provide useful guidance for users of the evidence and those who evaluate certainty in effects and associations at later time points. The ratings also help clarifying disagreement between organizations tackling similar questions about the evidence. Using the structured GRADE domains, narrative or other summaries (ideally in evidence profiles or summary of findings

EPA-HQ-2018-0008760045889

tables with explanations for the rating as shown in Table 1) of the evidence can be presented transparently with a judgment of the certainty in this evidence (Guyatt et al., 2011a; Langendam et al., 2016; Santesso et al., 2016).

Disclaimer

The content of the commentary should not be used for decision-making and represents a guide. It has not been officially endorsed by the GRADE Working Group but using GRADE without full systematic reviews although not ideal fulfills the criteria for using GRADE.

Conflict of interest

The authors declare they have no financial interests with respect to this manuscript, or its content, or subject matter. They are both members of the GRADE Working Group (www.gradeworkinggroup.org).

References

Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., et al., 2011. GRADE guidelines: 3. Rating the quality of evidence. J. Clin. Epidemiol.

Burford, B.J., Rehfuess, E., Schunemann, H.J., Akl, E.A., Waters, E., Armstrong, R., et al., 2012. Assessing evidence in public health: the added value of GRADE. J. Public Health (Oxf.) 34 (4), 631–635.

EFSA, 2007. EFSA Provisional Statement on a Request Form the European Commission Related to Melamine and Structurally Related to Compounds Such as Cyanuric Acid in Protein-Rich Ingredients Used for Food and Feed.

EFSA, 2008. Statement of EFSA on risks for public health due to the presences of melamine in infant milk and other milk products in China. EFSA J. 807, 1–10.

EPA, 2016. [Available from: https://www.epa.gov/flint.

GRADE_Working_Group. [Available from: http://www.gradeworkinggroup.org/intro.htm-criteria

Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., et al., 2011a. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J. Clin. Epidemiol. 64 (4), 383–394.

Guyatt, G.H., Oxman, A.D., Kunz, R., Brozek, J., Alonso-Coello, P., Rind, D., et al., 2011c. GRADE guidelines 6. Rating the quality of evidence–imprecision. J. Clin. Epidemiol. 64 (12), 1283–1293.

Guyatt, G.H., Oxman, A.D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., et al., 2011b. GRADE guidelines: 8. Rating the quality of evidence–indirectness. J. Clin. Epidemiol. 64 (12), 1303–1310.

Guyatt, G.H., Oxman, A.D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., et al., 2011d. GRADE guidelines: 7. rating the quality of evidence–inconsistency. J. Clin. Epidemiol. 64 (12), 1294–1302.

Guyatt, G.H., Oxman, A.D., Montori, V., Vist, G., Kunz, R., Brozek, J., et al., 2011e. GRADE guidelines: 5. Rating the quality of evidence–publication bias. J. Clin. Epidemiol. 64 (12), 1277–1282.

Guyatt, G.H., Oxman, A.D., Sultan, S., Glasziou, P., Akl, E.A., Alonso-Coello, P., et al., 2011f. GRADE guidelines: 9. Rating up the quality of evidence. J. Clin. Epidemiol. 64 (12), 1311–1316.

Guyatt, G.H., Oxman, A.D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., et al., 2011g. GRADE guidelines: 4. Rating the quality of evidence-study limitations (risk of bias) and publication bias. J. Clin. Epidemiol.

Harder, T., Takla, A., Rehfuess, E., Sanchez-Vivar, A., Matysiak-Klose, D., Eckmanns, T., et al., 2014. Evidence-based decision-making in infectious diseases epidemiology, prevention and control: matching research questions to study designs and quality appraisal tools. BMC Med. Res. Methodol. 14, 69.

Iorio, A., Spencer, F.A., Falavigna, M., Alba, C., Lang, E., Burnand, B., et al., 2015. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 350, h870.

Johnson, P.I., Sutton, P., Atchley, D.S., Koustas, E., Lam, J., Sen, S., et al., 2014. The navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122 (10), 1028–1039.

Lam, J., Koustas, E., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., et al., 2014. The navigation Guide — evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122 (10), 1040–1051.

Langendam, M., Carrasco-Labra, A., Santesso, N., Mustafa, R.A., Brignardello-Petersen, R., Ventresca, M., et al., 2016. Improving GRADE evidence tables part 2: a systematic survey of explanatory notes shows more guidance is needed. J. Clin. Epidemiol.

Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., et al., 2016. GRADE: assessing the quality of evidence in environmental and occupational health. Environ. Int. 92–93, 611–616.

National Toxicology Program, 2014. National Toxicology Program research project: West Virginia chemical spill. Board of Scientific Counselors December 9–10, 2014 Meeting ([Available from: http://ntp.niehs.nih.gov/go/974).

Neumann, I., Santesso, N., Akl, E.A., Rind, D.M., Vandvik, P.O., Alonso-Coello, P., et al., 2016. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. J. Clin. Epidemiol.

OECD, 2007. Guidance document on the validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. [Available from: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282007%292&doclanguage=en;.

Rooney, A.A., Cooper, G.S., Jahnke, G.D., Lam, J., Morgan, R.L., Boyles, A.L., et al., 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. Environ. Int.

Santesso, N., Carrasco-Labra, A., Langendam, M., Brignardello-Petersen, R., Mustafa, R.A., Heus, P., et al., 2016. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. J. Clin. Epidemiol. 92–93, 617–629.

Schunemann, H.J., Moja, L., 2015. Reviews: rapid! rapid! rapid! ...and systematic. Syst. Rev. 4 (1), 4.

Schunemann, H.J., Hill, S.R., Kakad, M., Bellamy, R., Uyeki, T.M., Hayden, F.G., et al., 2007a. WHO rapid advice guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. Lancet Infect. Dis. 7 (1), 21–31.

Schunemann, H.J., Hill, S.R., Kakad, M., Vist, G.E., Bellamy, R., Stockman, L., et al., 2007b. Transparent development of the WHO rapid advice guidelines. PLoS Med. 4 (5), e119.

Schünemann, H.J., Tugwell, P., Reeves, B.C., Akl, E.A., Santesso, N., Spencer, F.A., et al., 2013. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Res. Synth. Methods 4 (1), 49–62.

Spencer, F.A., Iorio, A., You, J., Murad, M.H., Schunemann, H.J., Vandvik, P.O., et al., 2012. Uncertainties in baseline risk estimates and confidence in treatment effects. BMJ 345, e7401.

U.S. Department of Health and Human Services. Contaminated water in flint 2016 [Available from: http://www.phe.gov/emergency/events/Flint/Pages/default.aspx

EPA-HQ-2018-0008760045890